

基于语义的大规模出租车轨迹交互分析方法

SemanticTraj: A New Approach to Interacting with Massive Taxi Trajectories

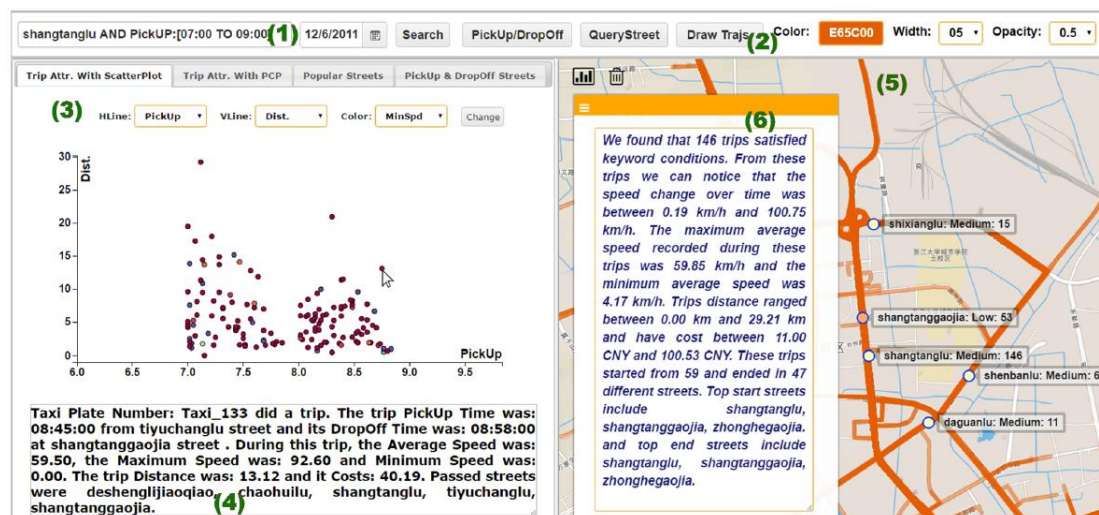
Shamal AL-Dohuki, Farah Kamw, Ye Zhao, Chao Ma, Yingyu Wu, Jing Yang, Xinyue Ye, Fei Wang, Xin Li, Wei Chen

简介:

在城市规划、城市运输及城市交通管理等领域，对于车辆轨迹的信息挖掘和知识提取越来越重要。大规模的城市出租车轨迹数据，能够帮助各个领域的专家分析城市交通和城市人群的移动规律。现有的工作大都围绕对地理位置的框选或刷选来完成数据的过滤。本文介绍了一种基于语义信息对大规模出租车轨迹数据进行交互式的检索、分析的可视化方法。领域专家或者使用者可以通过输入带有语义的查询条件对轨迹数据进行查询检索，并通过系统的可视化界面对轨迹数据进行探索，从而得到有用的信息。系统通过在轨迹数据上建立了两种文本语义索引文件来加速语义查询速率，并设计了多视图的可视化界面来帮助用户分析轨迹数据。最后通过用户的实验验证了系统的可行性。

问题导向:

大规模的出租车数据对城市交通城市规划等领域提供了很多有用的信息。为了挖掘出这些信息，现有的大量工作研究了轨迹信息的可视分析，轨迹信息的存储和轨迹语义的标注。然而这些工作大都是基于对地理位置的刷选来完成对数据的过滤。对于特定的任务，例如一家商场的管理者，希望知道在城市的什么地点放置摆渡车辆，能够帮助人们来商场购物？交通警察希望知道哪些路段是经常堵车的路段，那些路段交通状况很好？如果使用原有的系统来回答类似上述的问题，操作就会十分繁琐。而如果可以直接通过语义来查询满足条件的出租车轨迹信息，那么分析的时间就会大大下降了。



界面设计:

系统的主界面如上图所示:

1 查询输入框，允许用户进行文本的输入条件，同时也可以打开查询条件的选择器（如下图）。查询支持布尔查询，模糊查询，渐变查询和范围查询。

Query Form

Boolean Query

Proximity Query

Fuzzy Query

Please select the time period and input street names. You can add multiple streets by "Add Name". You can choose necessary boolean operators for querying multiple streets.

Time Period:

12:00

TO

15:30

Street Names:

gushanlu

Add Name

AND

nanshanlu

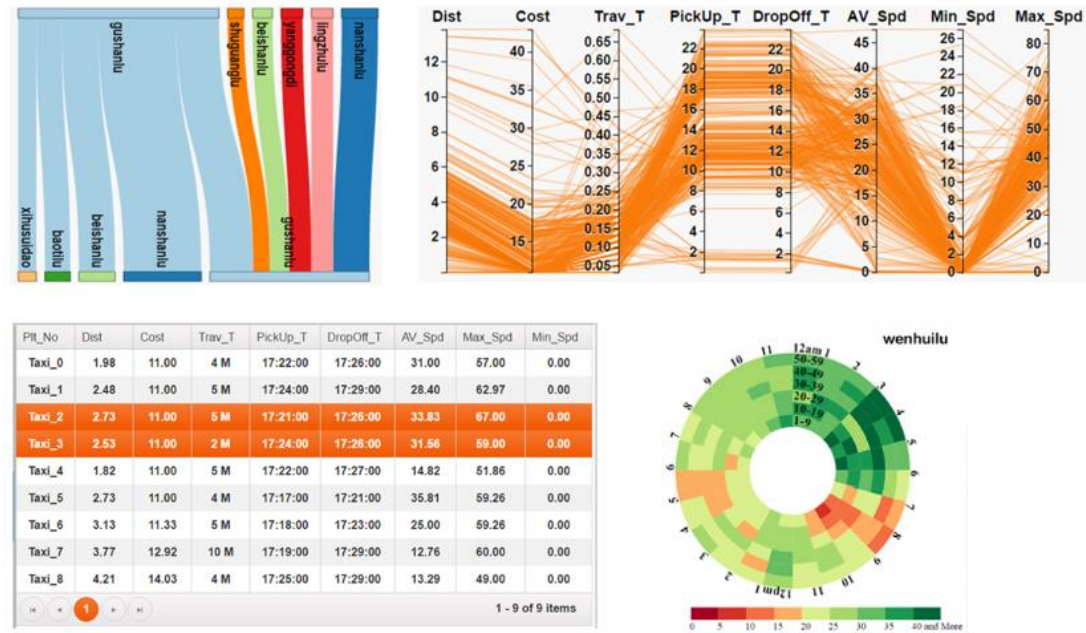
Remove

Generate Query

Cancel

2 绘制选项，系统提供了绘制选项使得用户可以自定义轨迹绘制的透明度，颜色宽度等属性，完成自定义的可视化效果。

3、4、5 轨迹的各种分析视图，包括反应道路交通流向关系的桑基图，轨迹数据各个维度的平行坐标图，数据的详细列表，和反应轨迹数据速度的环形热力图。并用文本的方式总结了轨迹的基本特点，包括轨迹的速度，车辆数目，上下客的最多的 5 条道路等信息。

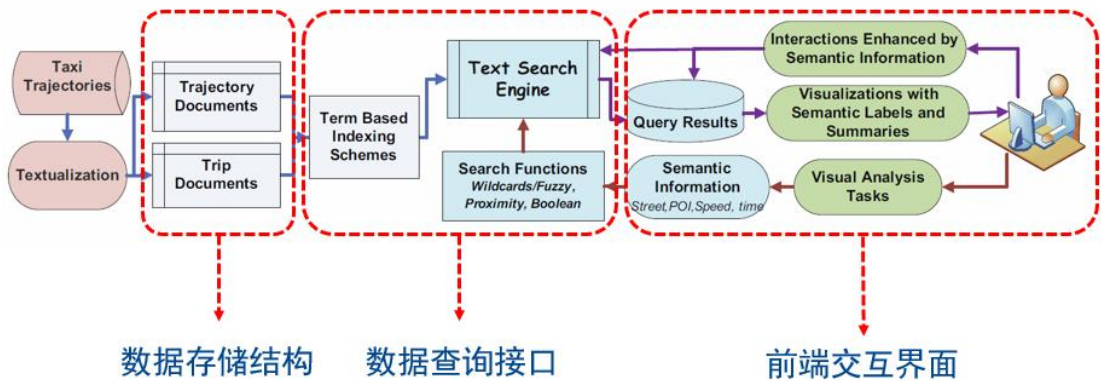


6 地图面板，用于在地图上直观的绘制轨迹，并添加文本的标签帮助用户理解轨迹的语义信息。

数据处理方法：

ID	经度	纬度	时间戳	速度	方向	载客状态
Taxi_1	120.2568	28.5689	14/1/10 12:25	45km/h	3	有客
Taxi_1	120.2567	28.5689	14/1/10 12:26	45km/h	3	有客
Taxi_1	120.2567	28.5689	14/1/10 12:27	20km/h	3	有客
Taxi_2	120.2358	28.5891	14/1/10 12:25	60km/h	1	空载

出租车的轨迹数据如上图所示，记录了某城市一个月的 8120 辆出租车的轨迹信息。基本的数据处理流程如下图。



系统通过文本语义的转换工具，把原始的出租车轨迹数据转换成两种不同类型的数据索引文件，包括出租车轨迹索引文件和行程索引文件。前者记录某出租车的轨迹语义索引，后者通过上下客信息记录某一段乘客的行程语义信息。并通过行驶距离计算出行驶的花费。

基于一条轨迹的数据存储

基于一段行程的数据存储

- 车辆ID
- 街道名称
- 速度信息
- 是否有客
- GPS轨迹

Index of a trajectory document of a taxi in a given time period

	Taxi Plate Number					
Street Names	Street1	Street1	Street2	Street2	...	
Speed	18.2	13.4	70.3	110.1	...	
DSpeed	Slow	Slow	Fast	Very Fast	...	
Empty or Not	Y	N	N	N	...	
GPS	22.533 114.044	22.533 114.046	22.532 114.049	22.532 114.050	...	

Index of a trip document of a taxi

	Taxi Plate Number				
	Pick-up Street		Pick-up Time		
	Drop-off Street		Drop-off Time		
	Travel Distance		Fare		
Street Names	S1	S2	S2	S4	...
GPS	22.533 114.044	22.533 114.046	22.532 114.049	22.532 114.050	...

- 车辆ID
- 上客街道，时间
- 下客街道，时间
- 行路路程，费用
- GPS轨迹

转换过程中，每一条数据的街道名称通过距离数据经纬度最近的街道来获取，而速度信息也通过赋予速度数值语义的方式转化为用户更能理解的数据格式，方便用户进行语义查询。

分析及案例：

系统通过了交通专家，犯罪学家及 15 名计算机学生的实验，通过该实验，结果表明轨迹数据的语义信息有利于更快速有效的理解数据。系统展示的各种统计图表和查询效率非常有利于交通问题的分析。系统中总结的文本信息有利于对数据信息的挖掘。证明了系统的有效性。

查询速率方面，作者提出了三个问题：

- Q1: 查询经过上塘路的所有出租车轨迹
- Q1: 查询经过上塘路和中河高架的所有出租车轨迹
- Q1: 查询经过上塘路或中河高架的所有出租车轨迹

验证系统的查询速率。结果表明所有的查询都能控制在交互级别的时间。证明了语义索引文件的有效性。

Table 2. Query performance on trip documents.

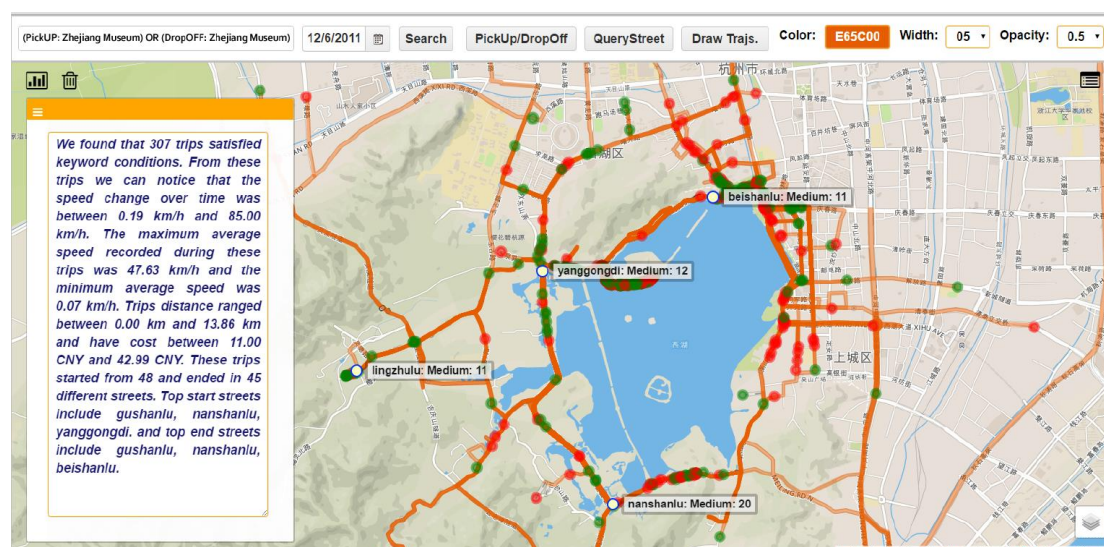
Time Period	Index Size	Indexing Time (sec)	Q1 Time (sec)	Q1 Hits	Q2 Time (sec)	Q2 Hits	Q3 Time (sec)	Q3 Hits
One Day	297MB	9	0.15	25k	0.14	6k	0.19	39k
One Week	1.00GB	336	0.21	90k	0.19	19k	0.32	137k
One Month	6.55GB	2,012	1.75	590k	1.32	141k	2.85	918k

Table 3. Query performance on trajectory documents.

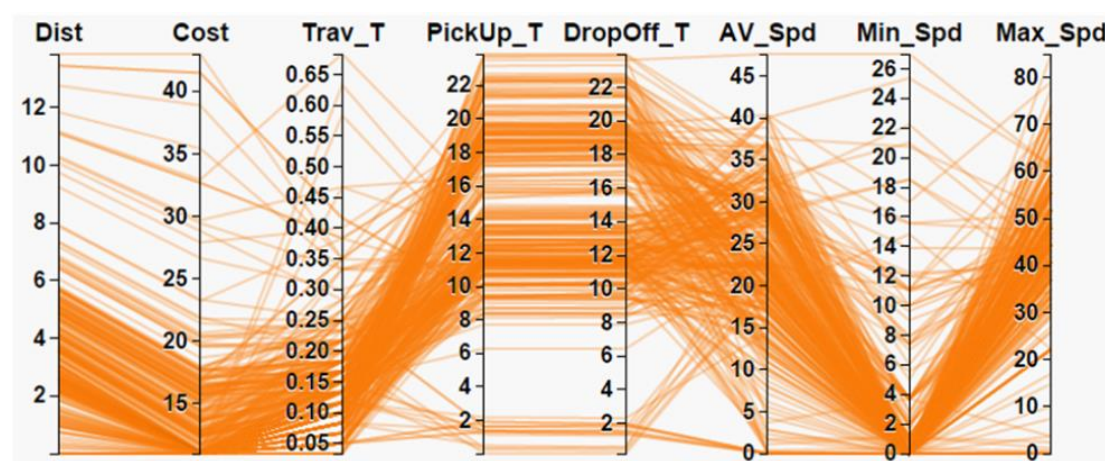
Time Period	No. of Files	Index Size	Indexing Time (sec)	Q1 Time (sec)	Q1 Hits	Q2 Time (sec)	Q2 Hits	Q3 Time (sec)	Q3 Hits
One Day	144	794MB	77	0.17	38k	0.21	3k	0.28	63k
One Week	1008	4.98GB	840	0.89	257k	0.65	19k	1.0	427k
One Month	4464	23GB	3,951	3.66	1,169k	2.14	84k	4.69	1,921k

最后通过案例分析，验证了系统在解决实际问题方面的有效性（举 1 例）：

浙江博物馆的管理人员想要选择接送游客的班车地点。他们通过输入浙江博物馆作为上下客地点为条件，过滤出的出租车图例如下。绿色(红色) 点表示出租车的上客(下客)地点。用户通过总结的文本信息和绿色点的位置分布，很容易的找到来博物馆的人最多的街道名称和地点，从而能够建立游客接送点。



然后管理人员通过平行坐标图，发现，下午 3 点到 5 点的车辆明显比其他白天的车辆少很多，从而能够得出在下午 3 点和 5 点可以适当的减少车辆来节约成本的结论。



总结

语义信息和地理信息的结合将愈演愈热，就像计算机与人，缺一不可。

论文：

Al-Dohuki S, Kamw F, Zhao Y, et al. SemanticTraj: A New Approach to Interacting with Massive Taxi Trajectories.[J]. IEEE Transactions on Visualization & Computer Graphics, 2016:1-1.